

**REGRESI NONPARAMETRIK DENGAN PENDEKATAN DERET
FOURIER PADA DATA DEBIT AIR SUNGAI CITARUM**

**Intaniah Ratna Nur Wisisono¹, Ade Irma Nurwahidah², Yudhie Andriyana³,
Neneng Sunengsih⁴**

Departemen Statistika, Fakultas MIPA, Universitas Padjadjaran
Jl. Dipati Ukur No 35, Bandung, Jawa Barat, Indonesia^{1,2,3,4}

rnwintaniah@gmail.com¹, adeirmanurwahidah@gmail.com², y.andriyana@unpad.ac.id³,
neneng@unpad.ac.id⁴

DOI:<https://doi.org/10.15642/mantik.2018.4.2.75-82>

Abstrak

Debit air sungai adalah salah satu faktor yang mempengaruhi terjadinya banjir. Besaran debit air bervariasi dari waktu ke waktu sehingga dibutuhkan pemodelan untuk mengetahui resiko banjir. Analisis yang sering digunakan untuk pemodelan adalah analisis regresi. Pemodelan regresi dapat dilakukan dengan tiga pendekatan yaitu pendekatan parametrik, semiparametrik, dan nonparametrik. Pemodelan parametrik data debit air bisa menggunakan ARIMA Box Jenkins. Salah satu pendekatan nonparametrik yang dikembangkan adalah menggunakan deret Fourier. Regresi nonparametrik deret fourier menghasilkan kurva sinus cosinus, sehingga sebaran data yang berulang sangat sesuai didekati menggunakan deret fourier. Estimasi deret fourier dapat menggunakan OLS (*Ordinary Least Square*). Dalam regresi nonparametrik deret fourier tingkat kemulusan fungsinya ditentukan oleh *bandwidthnya* (K). Penentuan bandwidth optimal dapat menggunakan metode GCV (*Generalized Cross Validation*). Dari hasil penghitungan didapat jumlah K optimal adalah 16. Adapun *R-square* yang dihasilkan adalah 0.7295 yang berarti bahwa 72,95% total variansi dalam variabel Y (debit) dapat dijelaskan oleh model regresi nonparametrik deret fourier. Model regresi deret fourier memberikan nilai RMSE sebesar 50,51 lebih kecil dibanding nilai RMSE ARIMA(1,0,0) yaitu 83,10 sehingga dapat disimpulkan bahwa regresi nonparametrik deret fourier lebih baik dalam memodelkan debit air sungai Citarum.

Kata kunci: Fourier, ARIMA, Debit Air, Nonparametrik

Abstract

River discharge is one of the factors that affect the occurrence of floods. It varies over time and hence we need to predict the flood risk. Since the plot of the data changes periodically showing a sines and cosines pattern, a nonparametric technique using Fourier series approach may be interesting to be applied. Fourier series can be estimated using OLS (*Ordinary Least Square*). In a Fourier series, nonparametric regression the level of subtlety of its function is determined by their bandwidth (K). Optimal bandwidth determined using the GCV (*Generalized Cross Validation*) method. From the calculation results, we have optimal bandwidth which is equal to 16 with R^2 is 0.7295 which means that 72.95% of the total variance in the river discharge variable can be explained by the Fourier series nonparametric regression model. Comparing to a classical time series technique, ARIMA Box Jenkins, we obtained ARIMA (1,0,0) with RMSE 83.10 while using Fourier series approach generate a smaller RMSE 50.51.

Keyword: Fourier, ARIMA, River discharge, Nonparametric

1. Pendahuluan

Citarum adalah sungai terbesar dan terpanjang di provinsi Jawa Barat. Sungai ini sangat mempengaruhi kehidupan masyarakat disekitarnya. Pemanfaatan sungai Citarum sangat bervariasi dari hulu hingga hilir baik untuk memenuhi kebutuhan rumah tangga, irigasi, pertanian, peternakan maupun Industri.

Banjir di Kabupaten Bandung adalah sejarah yang terus berulang. Sejak puluhan tahun bahkan ratusan tahun lalu, kawasan hulu dari sungai Citarum ini terus mengalami banjir. Derasnya luapan sungai Citarum sulit untuk dibendung. Meskipun pernah ada proyek besar untuk normalisasi Citarum dengan memotong aliran sungai yang berkelok dan mengalirkannya ke waduk Jatiluhur di Purwakarta, namun tetap saja ketika hujan deras, Citarum pasti meluap dan memicu banjir di Kabupaten Bandung.

Salah satu faktor yang mempengaruhi terjadinya banjir adalah debit air sungai. Debit air sungai adalah tinggi permukaan air sungai yang terukur oleh alat ukur permukaan air sungai [5]. Debit air sungai dalam hidrologi adalah tinggi permukaan air sungai yang terukur oleh alat ukur permukaan air sungai. Dalam sistem satuan SI besaran debit dinyatakan dalam satuan meter kubik per detik (m^3/detik).

Salah satu analisis statistik yang digunakan untuk pemodelan adalah analisis regresi. Pendekatan regresi dapat dilakukan dengan tiga pendekatan yaitu pendekatan parametrik, semiparametrik, dan non-parametrik. Data debit air sungai adalah data deret waktu karena diukur dari waktu ke waktu (dalam hal ini setiap bulan) sehingga dalam analisisnya perlu digunakan metode analisis untuk data deret waktu. ARIMA Box Jenkins adalah metode analisis data deret waktu parametrik yang paling sering digunakan [11]. Salah satu pendekatan lain, pendekatan nonparametrik yang dapat digunakan dalam analisis data deret waktu adalah deret fourier.

Deret fourier adalah deret yang digunakan dalam bidang rekayasa. Deret ini pertama kali ditemukan oleh seorang ilmuwan Perancis Jean-Baptiste Joseph Fourier (1768-1830). Deret ini juga dikenal sebagai deret

dalam bentuk sinus dan cosinus yang digunakan untuk mem-presentasikan fungsi-fungsi periodik secara umum sehingga sebaran data yang berulang/periodik sebagaimana terlihat pada pola data debit air sungai Citarum sangat sesuai didekati menggunakan deret ini.

Penelitian sebelumnya tentang pendekatan deret fourier dalam analisis data deret waktu nonparametrik antara lain; Model Regresi Nonparametrik Dengan Pendekatan Deret Fourier Pada Pola Data Curah Hujan di Kota Semarang oleh Nurjanah dkk [6], Model Regresi Nonparametrik Dengan Pendekatan Deret Fourier Pada Kasus Tingkat Pengangguran Terbuka Di Jawa Timur oleh Prahutama [8], kemudian Tripena (2007) mengkaji estimator deret Fourier pada regresi nonparametrik [9]. Sedangkan untuk regresi semi-parametrik menggunakan deret fourier telah dikembangkan oleh Asrini [1].

Estimasi deret fourier dapat menggunakan OLS (*Ordinary Least Square*). Dalam regresi nonparametrik deret fourier tingkat kemulusan fungsinya ditentukan oleh *bandwidth*-nya. Penentuan *bandwidth* optimal dapat menggunakan metode GCV (*Generalized Cross Validation*) [10]. *Bandwidth* yang optimal akan memberikan kurva yang mulus, variasi yang rendah dan bias yang besar. Dari nilai GCV yang minimum maka akan didapatkan *bandwidth* (K) yang optimal. Penelitian ini memodelkan debit air sungai Citarum menggunakan regresi nonparametrik dengan pendekatan deret fourier dan akan dibandingkan dengan hasil pemodelan ARIMA Box Jenkins.

2. Tinjauan Pustaka

2.1 Metode Regresi

Metode regresi adalah suatu metode statistik yang sering kali digunakan untuk menyelidiki dan memodelkan hubungan antara variabel respon Y dan variabel prediktor X . Misalnya diberikan himpunan data $\{(X_i, Y_i)\}, i = 1, \dots, n$ Secara umum hubungan antara Y dan X dapat ditulis sebagai berikut:

$$Y_i = m(X_i) + \varepsilon_i \quad (1)$$

dengan $m(X_i)$ merupakan suatu fungsi regresi yang belum diketahui dan akan ditaksir, dan ε_i adalah suatu variabel acak yang menggambarkan variasi Y di sekitar $m(x)$.

Penaksiran fungsi regresi dapat dilakukan dengan dua cara yaitu secara parametrik dan nonparametrik. Pada regresi parametrik digunakan bentuk fungsi parametrik tertentu sebagai $m(x)$. Fungsi $m(x)$ digambarkan oleh sejumlah parameter yang harus ditaksir. Dalam regresi parametrik terdapat beberapa asumsi terkait dengan model, sehingga diperlukan pemeriksaan terhadap asumsi-asumsi tersebut.

Pada regresi nonparametrik, fungsi regresi $m(x)$ ditaksir tanpa referensi bentuk kurva tertentu. Cara ini lebih fleksibel karena tidak memerlukan informasi apapun tentang sebaran data. Dengan suatu mekanisme matematis tertentu $m(x)$ akan mengikuti bentuk data.

2.2 Regresi Nonparametrik Deret Fourier

Salah satu model regresi dengan pendekatan nonparametrik yang dapat digunakan untuk menduga kurva regresi adalah regresi nonparametrik deret fourier. Deret fourier merupakan polinomial trigonometri yang mempunyai fleksibilitas, sehingga dapat menyesuaikan diri secara efektif terhadap sifat lokal data. Deret fourier baik digunakan untuk menjelaskan kurva yang menunjukkan gelombang sinus dan cosinus.

Diberikan data berpasangan sebagai berikut (y_j, t_{ij}) dengan $j = 1, 2, 3, \dots, n$ menyatakan banyaknya pengamatan dan $i = 1, 2, 3, \dots, p$ menyatakan banyaknya variabel bebas. Karena hanya terdapat satu variabel bebas, model regresi nonparametriknya sebagai berikut:

$$y_j = X_j \beta + \varepsilon \quad (2)$$

Jika $y = f(t) + \varepsilon$ maka $f(t) = [f(t_1)f(t_2)f(t_3) \dots f(t_n)]$
 $f(t)$ merupakan kurva yang tidak diketahui bentuknya maka $f(t)$ didekati dengan menggunakan deret fourier. Sebelum

pembentukan model fourier, variabel t ditransformasi terlebih dahulu menjadi:

$$t = 2\pi(t - \min(t))/(\max(t) - \min(t)) \quad (3)$$

Dengan model fourier sebagai berikut [2]:

$$y = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \sum_{k=1}^K \lambda_k \sin(tk) + \delta_k \cos(kt) + \varepsilon \quad (4)$$

2.3 Pemilihan Bandwith Dalam Regresi Nonparametrik Deret Fourier

Pemilihan bandwith sangat penting, karena berpengaruh pada model regresi nonparametrik deret fourier yang akan dipilih. Ada 2 strategi untuk memilih bandwith yang baik. Strategi pertama adalah memilih banyaknya bandwith yang relatif sedikit, sedangkan strategi yang kedua adalah sebaliknya, yakni menggunakan bandwith yang relatif banyak. Diantara kedua strategi tersebut, strategi kedua lebih banyak digunakan pada model yang sangat memperhatikan pola matematis yang ada pada data. Sedangkan strategi pertama, lebih mengarah pada alasan kesederhanaan model.

Pemilihan *bandwidth* yang terlalu kecil akan menghasilkan kurva yang *under-smoothing* yaitu sangat kasar dan sangat fluktuatif, dan sebaliknya *bandwidth* yang terlalu lebar akan menghasilkan kurva yang *over-smoothing* yaitu sangat mulus, tetapi tidak sesuai dengan pola data [4].

Penentuan lokasi *bandwith* yang berbeda akan menghasilkan model regresi nonparametrik deret fourier yang berbeda pula. Lokasi *bandwith* tersebut akan berpengaruh terhadap nilai kriteria dari model regresi nonparametrik deret fourier yang dibentuk.

Salah satu metode pemilihan titik bandwith yang optimal adalah *Generalized Cross Validation* (GCV). Model regresi nonparametrik deret fourier yang sesuai berkaitan dengan titik bandwith yang optimal didapat dari nilai GCV minimum. Fungsi GCV didefinisikan sebagai [10]:

$$GCV(K) = \frac{MSE(K)}{(n^{-1} \text{trace}(\mathbf{I} - \mathbf{A}))^2} \quad (5)$$

$$MSE(K) = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \quad (6)$$

dengan y_i = variabel respon
 \hat{y}_i = hasil dugaan dari y_i
 n = banyaknya pengamatan

2.4 Metode Box-Jenkins

Metode Box-Jenkins merupakan metode yang digunakan untuk analisis dan peramalan data kurva waktu (*time series*). Metode ini dapat digunakan pada data seret waktu yang stasioner dan terdiri dari tiga langkah yaitu identifikasi model, pendugaan parameter, dan diagnostik model [7].

Identifikasi model merupakan tahap untuk menentukan model-model sementara, yaitu dengan menentukan nilai p , q dan d . Penentuan nilai-nilai tersebut dilakukan dengan mengamati grafik fungsi ACF (korelogram) dan PACF (korelogram parsial). Nilai p (ordo proses AR) dapat ditentukan dengan melihat nilai pada grafik fungsi PACF dan nilai q (ordo proses MA) dapat ditentukan dengan melihat nilai pada grafik fungsi ACF.

Setelah identifikasi model tahap selanjutnya adalah pendugaan parameter. Pendugaan parameter bertujuan untuk menentukan apakah parameter sudah layak digunakan dalam model. Pendugaan parameter dapat dilakukan dengan menggunakan beberapa metode, yaitu metode momen, kuadrat terkecil dan kemungkinan maksimum (*likelihood*).

Tahap ketiga yaitu diagnostik model dilakukan untuk melihat model yang relevan dengan data. Pada tahap ini model harus dicek kelayakannya dengan melihat sifat sisaan dari sisi kenormalan dan kebebasannya. Secara umum pengecekan kebebasan sisaan model dapat dilakukan dengan menggunakan uji Q modifikasi Box-Pierce (*Ljung-Box*).

Selanjutnya yaitu pengecekan pada kenormalan sisaan dengan melakukan uji *shapiro-wilk normality*. Jika nilai- p yang dihasilkan $> \alpha$, maka dapat disimpulkan bahwa sisaan telah memenuhi asumsi kenormalan sisaan. Setelah semua proses dalam metode Box-Jenkins dilakukan

tahap berikutnya adalah melakukan *overfitting* model yaitu membandingkan model dengan model lain yang berbeda satu ordo di atasnya. Hal yang dibandingkan pada *overfitting* adalah signifikansi parameter, pemenuhan asumsi sisaan, dan *Akaike's Information Criterion* (AIC). Jika dalam proses *overfitting* didapatkan model yang relevan dengan data, maka langkah terakhir adalah proses peramalan. Peramalan merupakan proses untuk menentukan data beberapa periode waktu kedepan dari titik waktu ke- t .

2.5 Kriteria Model Terbaik

Akaike's Information Criterion (AIC) ditemukan oleh Akaike dimana kriteria ini dapat digunakan untuk melakukan perbandingan model pada data yang sama. AIC adalah metode yang digunakan untuk mendapatkan model terbaik. AIC didefinisikan sebagai:

$$AIC = -2l(y; \alpha) + 2 \text{trace}(H) \quad (6)$$

dimana $l(y; \alpha)$ adalah fungsi likelihood. Suatu model dikatakan model terbaik jika memiliki nilai AIC terkecil.

Sebagai tambahan, Eubank menyebutkan bahwa kinerja dari hasil pendugaan model regresi dapat dilihat dari nilai *Root Mean Square Error* (RMSE) dimana diformulasikan sebagai berikut [3]:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (7)$$

Selanjutnya untuk mengetahui kualitas suatu model, kita dapat menggunakan koefisien determinasi (R^2). Nilai R^2 menunjukkan kemampuan model dalam menjelaskan variabilitas data. Semakin tinggi nilai R^2 maka semakin baik kualitas dari suatu model. Nilai R^2 diperoleh dengan rumus:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (8)$$

3. Metodologi Penelitian

3.1 Sumber Data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder yang bersumber

dari BBWS Citarum (Balai Besar Wilayah Sungai Citarum) periode Januari 2009 – Desember 2016 yang merupakan data bulanan. Variabel respon (Y) yang digunakan dalam penelitian ini adalah debit air sungai tertinggi sedangkan variabel bebas (X) yang digunakan pada penelitian ini adalah waktu.

3.2 Tahapan Penelitian

Langkah – langkah analisis dalam penelitian ini adalah sebagai berikut :

1. Membuat plot *time series* debit air sungai
2. Memilih jumlah bandwidth yang optimal berdasarkan nilai dari GCV
3. Mendapatkan model regresi nonparametrik deret fourier dengan bandwidth yang optimum
4. Menghitung nilai R^2 , AIC, dan RMSE dari model regresi nonparametrik deret fourier
5. Melakukan pemodelan ARIMA Box Jenkins untuk data debit
6. Menghitung nilai RMSE dari model ARIMA Box Jenkins
7. Membandingkan nilai RMSE dari model regresi nonparametrik deret fourier dengan model ARIMA
8. Menarik kesimpulan

4. Hasil dan Pembahasan

4.1 Deskriptif Statistik

Berikut ini adalah deskriptif dari data debit air sungai

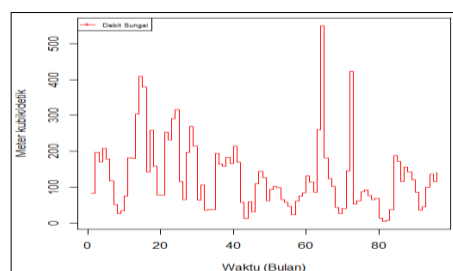
Tabel 1. Deskriptif Data	
Debit Air Sungai (m ³ /detik)	
Min.	5,25
Median	107,72
Mean	129,91
Max.	549,35
Stdev	97,62

berdasarkan tabel 1 diperoleh informasi bahwa debit air sungai memiliki nilai tertinggi sebesar 549,35 m³/detik dan nilai terendah sebesar 5,25 m³/detik. Dimana

nilai rata-ratanya adalah 129,91 m³/detik dan mediannya 107,72 m³/detik.

4.2 Plot Debit Air Sungai Terhadap Waktu

Pola hubungan antara waktu sebagai variabel prediktor terhadap debit air sungai sebagai variabel respon terlihat pada gambar 1. Dari gambar terlihat adanya pola periodik, dengan kata lain ada pola seperti sinus cosinus. Sehingga diharapkan regresi nonparametrik deret Fourier akan dapat memodelkan data debit ini dengan baik.



Gambar 1. Time Plot Debit Air Sungai

4.3 Pemodelan Regresi Nonparametrik Deret Fourier

Regresi nonparametrik deret fourier adalah salah satu metode yang digunakan untuk menaksir kurva regresi nonparametrik. Model nonparametrik deret fourier dengan titik bandwidths digunakan untuk menggambarkan pola periodik. Metode yang digunakan dalam menaksir parameter nonparametrik deret fourier adalah Metode Kuadrat Terkecil (*Ordinary Least Square*).

Kriteria yang harus diperhatikan dalam membentuk model regresi nonparametrik deret fourier adalah menentukan bandwidth (K) untuk model regresi. Nilai K merupakan bilangan bulat positif. Penentuan K optimal menggunakan kriteria GCV. Adapun hasil yang didapat dari setiap K yang dicobakan adalah sebagai berikut:

Tabel 2. Nilai GCV untuk setiap K

Nilai K	GCV	Nilai K	GCV
1	9429,19	11	7465,17
2	9214,37	12	7331,25
3	9618,46	13	7005,44
4	9162,12	14	7066,98
5	9538,69	15	6769,36
6	9996,96	16	6317,50
7	10340,79	17	6670,39
8	7742,79	18	7141,09
9	7578,16	19	7496,61
10	7961,14	20	8033,92

Berdasarkan tabel 2 terlihat nilai K optimum adalah 16 dengan nilai GCV sebesar 6317,50. Selain melihat nilai GCV dilihat juga nilai R^2 dan MSE untuk lebih meyakinkan bahwa K optimum adalah 16.

Tabel 3. Nilai R^2 dan MSE untuk setiap K

Nilai K	R^2	MSE
1	0,10	8472,56
2	0,16	7919,60
3	0,16	7919,60
4	0,24	7182,76
5	0,24	7182,76
6	0,24	7182,76
7	0,24	7182,76
8	0,47	4981,23
9	0,51	4625,34
10	0,51	4625,34
11	0,57	4083,32
12	0,60	3787,34
13	0,64	3412,26
14	0,64	3412,26
15	0,69	2915,32
16	0,73	2550,72
17	0,73	2550,72
18	0,73	2550,72

Berdasarkan nilai R^2 dan MSE untuk setiap K juga didapatkan Nilai K optimum 16 dengan R^2 sebesar 72,95% dan MSE sebesar 2550,72. Meskipun untuk $K > 16$ didapatkan nilai R^2 dan MSE yang sama dengan $K = 16$, dipilih K optimum 16 berdasarkan prinsip parsimoni, dipilih model yang lebih sederhana. Untuk nilai $K = 16$ ada 35 parameter yang harus diduga.

Setelah didapatkan nilai K optimum, selanjutnya dilakukan pendugaan parameter. Nilai dugaan parameter regresi nonparametrik deret fourier untuk data

debit sungai Citarum ditampilkan dalam tabel berikut:

Tabel 4. Nilai Dugaan Parameter (Statistik) untuk $K=16$

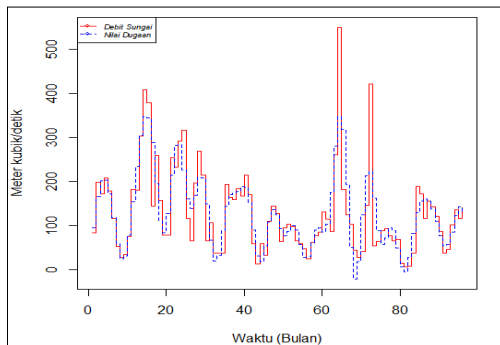
Statistik	Nilai	Statistik	Nilai
α_0	-829,61	λ_{16}	-26,95
α_1	905,30	δ_1	586,32
α_2	-143,19	δ_2	116,48
λ_1	45,03	δ_3	71,76
λ_2	36,31	δ_4	26,95
λ_3	5,67	δ_5	21,97
λ_4	-31,92	δ_6	23,17
λ_5	5,09	δ_7	14,14
λ_6	0,54	δ_8	52,61
λ_7	-8,48	δ_9	22,98
λ_8	51,82	δ_{10}	-2,76
λ_9	15,12	δ_{11}	-15,05
λ_{10}	6,63	δ_{12}	6,92
λ_{11}	28,05	δ_{13}	-4,55
λ_{12}	-24,56	δ_{14}	-1,50
λ_{13}	-27,05	δ_{15}	10,89
λ_{14}	18,78	δ_{16}	-5,78
λ_{15}	22,79		

Model regresi nonparametrik deret fourier untuk menggambarkan hubungan debit air sungai dan waktu dengan jumlah bandwidth 16 jika dituliskan persamaan regresinya adalah sebagai berikut:

$$\begin{aligned} \hat{f}(x) = & -829.61 + 905.3 t - 143.19 t^2 \\ & + 45.03 \sin t + 36.31 \sin 2t \\ & + 5.66 \sin 3t - 31.92 \sin 4t \\ & + 5.08 \sin 5t + 0.54 \sin 6t \\ & - 8.48 \sin 7t + 51.82 \sin 8t \\ & + 15.11 \sin 9t + 6.63 \sin 10t \\ & + 28.05 \sin 11t - 24.56 \sin 12t \\ & - 27.05 \sin 13t + 18.78 \sin 14t \\ & + 22.79 \sin 15t - 26.94 \sin 16t \\ & + 586.32 \cos t + 116.48 \cos 2t \\ & + 71.76 \cos 3t + 26.95 \cos 4t \\ & + 21.97 \cos 5t + 23.17 \cos 6t \\ & + 14.14 \cos 7t + 52.62 \cos 8t \\ & + 22.98 \cos 9t - 2.76 \cos 10t \\ & - 15.05 \cos 11t + 6.92 \cos 12t \\ & - 4.55 \cos 13t - 1.5 \cos 14t \\ & + 10.88 \cos 15t - 5.78 \cos 16t \end{aligned}$$

Untuk melihat seberapa baik model yang dibentuk dapat dengan membandingkan

pola plot antara data debit dengan nilai dugaannya.



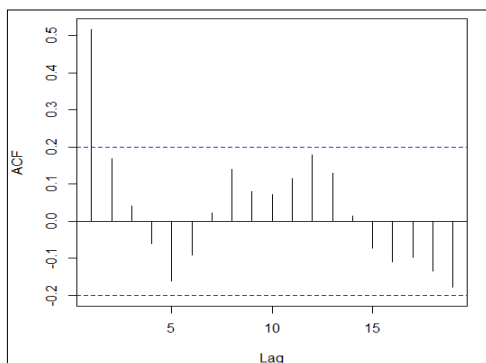
Gambar 2. Plot Model Regresi Nonparametrik Deret Fourier K 16

berdasarkan gambar 2 terlihat nilai dugaan dari model cukup menggambarkan pola dari data debit, meskipun ada beberapa titik yang tidak terjangkau.

Model regresi nonparametrik deret fourier dengan $K = 16$ menghasilkan model dengan $R^2 = 72,95\%$. Dapat diartikan bahwa 72,95% total variansi dalam variabel Y (debit) dapat dijelaskan oleh model regresi yang terbentuk.

4.4 Pemodelan ARIMA Box-Jenkins Referensi

Langkah pertama dalam pemodelan ARIMA Box Jenkins adalah pemeriksaan kestasioneran data. Kestasioneran data dapat dilihat dari pola ACF dan uji *Augmented Dickey Fuller*.



Gambar 3. Plot ACF Debit Air Sungai

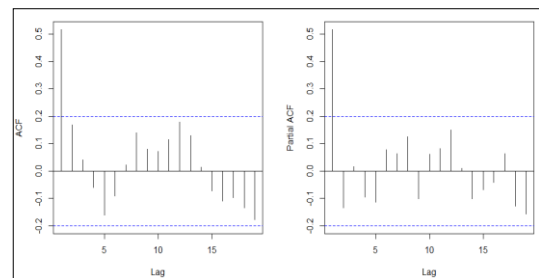
Berdasarkan gambar 3 terlihat pola plot ACF data terlihat pola *sinusoidal* serta hasil uji *Augmented Dickey Fuller*

memberikan p-value sebesar $0,01 < \alpha = 0,05$ dimana hipotesisnya adalah:

H_0 = Data tidak stasioner

H_1 = Data stasioner.

Sehingga disimpulkan data stasioner. Selanjutnya dilakukan identifikasi model ARIMA dari plot ACF dan PACF.



Gambar 4. Plot ACF dan PACF Debit Air Sungai

Berdasarkan gambar 4 dipilih kandidat model ARIMA(1,0,1), ARIMA(1,0,0), dan ARIMA(0,0,1). Dari ketiga kandidat model didapat nilai AIC untuk ARIMA(1,0,1) sebesar 1125,44 ARIMA(1,0,0) sebesar 1125,39 dan ARIMA(0,0,1) sebesar 1126,49. Karena nilai AIC model kedua lebih kecil maka model terbaik dari kandidat model adalah ARIMA(1,0,0).

4.5 Perbandingan Model

Perbandingan antara model regresi nonparametrik deret fourier dan model ARIMA menggunakan kriteria RMSE ditampilkan pada tabel 5.

Tabel 5. Perbandingan nilai RMSE

Model	RMSE
Deret Fourier	50,51
ARIMA(1,0,0)	83,10

Dari tabel 5 terlihat model regresi nonparametrik deret fourier memiliki nilai RMSE yang lebih kecil dibandingkan model ARIMA(1,0,0). Jadi model terbaik diantara kedua model tersebut adalah model regresi nonparametrik deret fourier.

5. Kesimpulan

- 1) Dengan menggunakan nilai GCV pada model regresi nonparametrik deret fourier diperoleh jumlah bandwidth yang optimum adalah 16.

- 2) Model regresi nonparametrik deret fourier yang digunakan untuk menggambarkan hubungan debit air sungai dan waktu pada penelitian ini adalah sebagai berikut:

$$\begin{aligned}\hat{f}(x) = & -829.61 + 905.3 t - 143.19 t^2 \\ & + 45.03 \sin t + 36.31 \sin 2t \\ & + 5.66 \sin 3t - 31.92 \sin 4t \\ & + 5.08 \sin 5t + 0.54 \sin 6t \\ & - 8.48 \sin 7t + 51.82 \sin 8t \\ & + 15.11 \sin 9t + 6.63 \sin 10t \\ & + 28.05 \sin 11t - 24.56 \sin 12t \\ & - 27.05 \sin 13t + 18.78 \sin 14t \\ & + 22.79 \sin 15t - 26.94 \sin 16t \\ & + 586.32 \cos t + 116.48 \cos 2t \\ & + 71.76 \cos 3t + 26.95 \cos 4t \\ & + 21.97 \cos 5t + 23.17 \cos 6t \\ & + 14.14 \cos 7t + 52.62 \cos 8t \\ & + 22.98 \cos 9t - 2.76 \cos 10t \\ & - 15.05 \cos 11t + 6.92 \cos 12t \\ & - 4.55 \cos 13t - 1.5 \cos 14t \\ & + 10.88 \cos 15t - 5.78 \cos 16t\end{aligned}$$

Dalam memodelkan data debit air sungai dalam penelitian ini, model regresi nonparametrik deret fourier lebih baik dibandingkan model ARIMA (1,0,0)

Referensi

- [1] Asrini, Luh Juni. “Regresi Parametrik Deret Fourier”, Prosiding Seminar Nasional FMIPA Universitas Negeri Surabaya, (2012) 77-80, 24 November, Surabaya.
- [2] Bilodeau, M. “Fourier Smoother and Additive Models”, The Canadian Journal of Statistics, 3, (1992) 257-259.
- [3] Eubank, R., Nonparametric Regression and Spline Smoothing. New York: Marcel Dekker. (1999).
- [4] Hardle, Wolfgang. 1994. Applied Nonparametric Regression. Springer-Verlag. Berlin. (1994).
- [5] Mulyana. Pemodelan Debit Air Sungai Studi Kasus DAS Cikapundung. Makalah, disampaikan pada Lokakarya Sistem Informasi Pengelolaan DAS: Inisiatif Pengembangan Infrastruktur Data. IPB. (2007). 5 September, Bogor.
- [6] Nurjanah, Fatmawati dkk. Model Regresi Nonparametrik Dengan Pendekatan Deret Fourier Pada Pola Data Curah Hujan di Kota Semarang. Universitas Muhammadiyah Semarang. Semarang. (2015).
- [7] Pankratz, A. Forecasting with Univariate Box-Jenkins Models: Concepts and Cases. John Wiley and Sons, New York. (1983).
- [8] Prahutama, Alan. Model Regresi Nonparametrik Dengan Pendekatan Deret Fourier Pada Kasus Tingkat Pengangguran Terbuka di Jawa Timur. Prosiding Seminar Nasional Statistika. Universitas Diponegoro. (2013). Semarang.
- [9] Tripena, A. Tesis. Estimator Deret Fourier dalam Regresi Nonparametrik, ITS, Surabaya. 2007.
- [10] Wu, H. dan Zhang, J.T. Nonparametric Regression Methods for Longitudinal Data Analysis, A John-Wiley and Sons Inc. Publication, New Jersey. (2006).
- [11] Ulinnuha, Nurissaidah dan Farida, Yuniar “Prediksi Cuaca Kota Surabaya Menggunakan Autoregressive Integrated Moving Average (ARIMA) Box Jenkins dan Kalman Filter”, Jurnal Matematika MANTIK, vol. 4, no. 1, hal. 59-67, Mei 2018